

Weekly Report 6.3 – 6.9

Done

1. Cctv 项目那边，对 tooltip 形式的讨论与实现（对于更好体现 tooltip 的办法，修改之前炳如写的相关代码逻辑）；对边缘发光功能的调试与改进（熟悉实现 svg 的滤镜功能，发现了改滤镜不能用于 path 和 line 的问题）
2. 新加坡小朋友们进行了 data journalism awards 的分享。他们和组内同学分享内容是不同的种类，但是比较凑巧的是新加坡同学似乎分配到的作品都更好看，大家也都很有想法；可能因为他们受到的思维约束更少。下周我将会让他们对这些作品他们认为好的、可用的地方整理；对他们未来想做的作品的数据类型（时序、事件、甚至时空）、形式，结合整理的优点进行初步头脑风暴。
 - a. 这次又出现了大量基于 unit visualization 的作品。它确实很适合用于讲故事，在保持逼格的同时不丢失表现力，并且不会太让人觉得腻。当然，这仍然是需要以良好的故事线和对数据的理解作为基础的。个人想法是，结合 unit visualization 的 design space, 结合半自动的故事线构建做一个生成工具。
3. 简单研究了一下 vis 文章的 review。大家的意见其实挺一致的，我个人也认为都很有道理。相比 vast, infovis 的 reviewer 更为严格、说的也十分直接。集中三点如下 abc。个人来看，b 是写作时就已经在考虑和讨论的，当初尽力引用了一些相关文献但是还不够；c 可以通过探访用户来做。
 - a. 我们做了从数据特征提取一直到视频编辑的所有过程，但其实这每一步都可以单独做一篇文章。
 - b. 序列生成算法需要说明为什么用了这些点，没用别的点？
 - c. 最后视频的质量没有评估；用户界面设计的选择没有评估。
4. 数据可视化创作大赛数据视频组的评选。我选中的最后都没有评上奖；但是有两个我认为作为单纯可视化作品不错、但是作为可视化视频不太行（需要更好的视频制作和讲故事技巧）的作品评上了奖。
5. 其余想法 1: 在智云的视频中, 有些视频是有多个可视化在一个页面内同时播放；有些则是前后两个视频衔接。以这两种组织方式为基，形成一种故事线的数据结构，探索如何生成这样的结构。这里用到一些知识点，比如 visual comparison, composite visualization 等都是老问题, 但是还没看到用于视频的。
6. 其余想法 2: (自动) annotation 生成技术。这个东西在静态可视化/storytelling vis 中已经有不少作品了；但是在视频中没见到，而且从我近来观察（比如智云比以及微信群等）这个有不少实际需求。Annotation 什么内容、怎么加（可视化呈现？用什么动画效果？） 、时间控制等都可以。
7. 其余想法 3 见下面论文阅读 2.
8. 从我当前（简单）认知来说，构建叙事结构也类似于是是一个不断找关系（关联）的过程，各种类型数据、各种数据的性质（如时序数据的时间维度），若能通过某一个角度找到关联，就可以放一起或者以其他形式衔接来体现这种关系。就像陈老师分享的 StoryMap 介绍文章来说，一般讲故事都拿时间来串联，而 storymap 是用空间信息来的。

- a. 这里又涉及一个人到底应该起多大作用的问题。Primary reviewer 很狠的裁定我们这种自动算法不可能生成 interesting 的结果；storymap 也是一个给定模板后面手动的工具。自动到底需要到何种程度？

To Do

1. 对论文修改的规划。
2. 继续整理、讨论近来的想法。
3. 新加坡和组内小朋友一起头脑风暴。
4. PVis 论文修改转投。

论文阅读:

1. VAST12 *Just-in-Time Annotation of Clusters, Outliers, and Trends in Point-based Data Visualizations* 本文提出的主要是这个 just-in-time 的概念(交互可视化的形式), 并以聚类离群和趋势作为驱动, 通过提取语义信息并标注. 同样, 也是二维空间发现视觉模式; 但是语义信息生成是从高维空间做的. 同样他也面临一些 quality evaluation 的问题, 但是巧妙的避开了.
2. CHI13 *Contextifier: Automatic Generation of Annotated Stock Visualizations* 这篇文章既结合了可视化上的视觉特征, 又加上了相关新闻报道信息, 做评分、排序与推荐. 在智云当前作品中, 只做了后者; 而我们做的, 可以认为是前者. 两者结合, 用于数据视频, 是否可以作为一个创新?
3. CHI19 *GameViews: Understanding and Supporting Data-driven Sports Storytelling* 根据篮球视频、比分信息、文字直播, 三种数据联合找到有趣的事件, 时间对齐并呈现. 尽管是叫 storytelling, 但其实本身还是一个交互系统, 有交互性. 形式很简洁.
4. CHI19 *VizML: A Machine Learning Approach to Visualization Recommendation* 从一些可视化网站上爬了可视化作品, 把数据集已经对于各种可视化的编码等信息拿下来用于训练, 变成一个推荐算法. 值得注意的是今年 CHI 还有一篇文章, 做了常见用于可视化的数据集还有评估可视化结果的 benchmark.
5. Medium 上的博文 *Let's Talk about Natural Language Interfaces (NLI) for Data Visualization* 可视化里为什么要用 NLI? 我本来以为只是为了偷懒(比如语音输入), 但是本文提出的原因是这样, 对于很多可视化系统, steep learning curve of tool + people have to translate the question in their minds into actions supported by the tool. 学起来难; 问题需要转化成为实际交互操作才能用; 这个转化还需要是系统支持的操作. 本文还提出一些现有可视化中 NLI 面对挑战, 如歧义、如信息过少、如上下文支持(之前的语言输入对后面的 NLP 解释也要有影响). 然后是一些未来方向如自然语言作为输出和用自然语言辅助可视化来回答问题. 其实在 09 年, 就有一篇对于视频故事构建利用文字输入的办法, 做一些预测.

规划:

短期: data2video 的想法整合、与福风的讨论。上周末完成, 这周继续。

短期: 新加坡小朋友的可视化学习以及两边一起的头脑风暴 (June 16) 。

中期: 本周在前端调试方面有进步。后面继续跟进, 灵活分配算法相关学习和 cctv 这边时间分配。

中期: vis 论文的修改

工作时间

平时 40, 假日 12 小时. 总共 52 小时.